DOCUMENT RESUME

ED 421 698 CS 216 392

AUTHOR Pounds, Buzz R.
TITLE Standardization, Diversity, and Teacher Evaluation of Writing.
PUB DATE 1998-04-02
NOTE 23p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (49th, Chicago, IL, April 1-4, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Freshmen; *Evaluation Criteria; Evaluation Methods; Evaluation Research; *Freshman Composition; Higher Education; Statistical Analysis; *Student Evaluation; Teacher Role; *Writing Evaluation; Writing Research; *Writing Teachers
IDENTIFIERS *Faculty Attitudes; University of Northern Iowa

ABSTRACT

A study examined whether or not the composition faculty at the University of Northern Iowa agreed on criteria for evaluating student writing and to what extent they would give similar scores even if their criteria did not agree. Using 15 sample student papers written for the Writing Competency Examination (WCE), the scores given by current writing faculty using a 4-point holistic scale in addition to one 3x5 index card's worth of comments were analyzed. Results indicated that scores reflected a strong tendency toward the middle range. Statistical analysis on the subgroups revealed that of several factors the most significant was the relationship of scores to whether or not the rater made more content or form comments, followed by percentage of positive comments and teaching classification. Findings suggest that since significant differences were not found among the subgroups, standards are not sacrificed in spite of the current practice which reflects a more diverse and possibly less experienced group of teachers. The study suggests that: (1) the current system produces scores which tend toward the average with few highs and fewer lows; (2) no standardization of criteria exists among teachers of the College Reading and Writing (CRW) course; and (3) the university is served by promoting academic freedom for all teachers of first-year composition. (Contains 7 tables of data and 28 references; various sample forms--a recruitment letter for the study, a demographic information questionnaire, and two additional tables of data--are appended.) (CR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
B. Pounds
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

ED 421 698

Buzz R. Pounds
CCCC 98
2 April 1998

Standardization, Diversity, and Teacher Evaluation of Writing

While service learning has protean definitions, this paper uses the definition mentioned earlier that first-year composition is mandated, as at least containing a skills element, to be a service course for the entire university. As such, the seriousness and applicability of this mandate require more than perhaps the support given by the university. Can the diversity of teachers and pedagogies fulfill the mandate to teach skills for the curriculum without either weakening standards or becoming too standardized? While the designers of large-scale assessments assumed that good writing was easily defined and applied, no such explicit assumption is now in place. As no study of how the different instructors (Graduate Assistants, Adjuncts, and Tenure-Track Professors) approach student papers has been done, I undertook this study to ascertain how the different teachers of first-year composition scored essays and by what criteria they evaluated them. This study evaluates the current status of first-year composition and the ability of these different teachers to agree on a standard of writing itself to see what is gained and lost by the diversity of first-year composition instructors, and the implications for serving the university.

First a brief historical overview. The University of Northern Iowa has come full circle. The change was from two classes which emphasized form to an assessment, the former Writing Competency Examination (WCE), which also emphasized form to a class, Introduction to College Reading and Writing (CRW). As the change developed, UNI writing faculty went from a few writing specialists to several transitory graduate assistants, various adjuncts, and

CS216392

percentage of the faculty--composition and literary specialists alike. When writing was formalistic, the writing staff attained agreement; now, does the diversity of teachers and their diversity of criteria place agreement out of reach? Underlying the WCE was the assumption that

> by collective professional judgment. . . . [t]he department of English Language and Literature can define the criteria of basic writing competency. . . . provide instruction in these skills as needed. . . . evaluate samples of writing in a consistent and reliable way in order to certify the demonstration of basic writing competency. (Senate 24 Jan. 1977 15)

Nationwide, assessments, while still practiced, began to decline due to the understanding that writing well involved more than the ability to produce one essay for one situation, both for writers in general and students in particular. The outcome of better writing is related to the ability and usefulness of testing writing. Whether positive or negative, backwash is the effect a test or assessment has on students, teachers, and/or the institution (cf. Hughes 1-2). Students' writing improvement is positive backwash; students' dislike for writing is negative backwash. Backwash also affects the curriculum. Marie Jean Lederman notes that in spite of the long history of testing "we continue to worry whether or not the format of an essay examination will have a negative affect on students' creativity and thinking or, worse, that our tests may become more important than our curriculum" (37).

First-year composition developed as a backwash effect of the WCE as the WCE did not consistently facilitate appropriate placement in the sequence of the university's curriculum. The test did not serve the greater curriculum as student ability did not transfer to upper-division course work due to students taking the course late in their academic careers, along with poor attitudes toward writing in general. Thus, the University Writing Committee promoted the return to a first-year Composition class in addition to a Writing-Across-the-Curriculum (WAC) program. The writing class emerged along with the new general education requirements, while

WAC was never fully implemented.

Before implementing the new General Education requirement, several issues were addressed by the faculty senate, and as such still concern the university. First, senators observed the question of funding as it impacts students' ability to make progress; that is, are there enough sections to keep from creating a backlog of upper division students needing to take these courses so that first-year students are unable to take the classes. Secondly, senators questioned who would teach these courses--tenure-track faculty, or adjuncts (with MAs).[1] Some diplomatically noted that, while not necessarily a problem (contrary to others who suggested these MA instructors would "water-down and weaken the general education program"), the creation of MA instructors may not produce quality teaching, nor would the creation of non-tenure track term employees be compatible with the university's structure (Senate 16 Nov. 1987 3, 5). The provost at the time promised funds for whichever plan promoted the best for education, but those funds never materialized and no new faculty were hired to teach the required 40 or more sections of CRW (Eblen).

As with many general education classes, CRW is taught by different faculty with different backgrounds, but most teachers expect the class to improve student writing by having the students write. In order for the course to achieve this goal across the many classes, the course should address issues of validity, reliability, and for these purposes, teacher qualification The lack of certainty in whether or not student learning can be documented may not be without a price. W. Ann Reynolds (at the height of the testing movement) describes the tension between faculty concerns for academic freedom and the accountability requested from students, trustees, and politicians (4-5) and proposes that the testing of teachers (in more than content areas) and students serve as only one small portion of obtaining greater achievement (7). John Chandler

forewarns that autonomy in the classroom (as traditionally expected and practiced), while providing intellectual outcomes, does not always produce specific curricular outcomes (12). Chandler writes, "the major national reports on the improvement of undergraduate education . . . charge that curricular incoherence is the result largely of the radical freedom of faculty members to teach what they like with little reference to the needs of the students" (12). Chandler acknowledges some difficulties with testing intellectual outcomes, but suggests that "the assessment movement holds considerable promise for encouraging faculties to exercise collective responsibility and to approach their educational tasks with a collegial mind-set" (13). Chandler also acknowledges the importance of keeping testing under the control of faculties, but notes that "to be credible and effective in the exercise of their responsibility for assessment, it is imperative that faculty members surrender some of their individual autonomy and work collaboratively" (15). To get a sense of whether diversity, predicated on a commitment to academic freedom, promotes lack of agreement, how widespread would scores be?

Research Situation

At UNI, tenure-track professors (hereafter "professors") rotate general education assignments, roughly one class every four semesters. The current English faculty include 53 teachers (33 professors, 9 adjuncts, and 11 teaching graduate assistants) who teach a variety of General Education courses in addition to CRW. GAs and Adjuncts usually have the opportunity to teach only CRW; some GAs also assist large sections of Humanities (Table 1). In one sample semester, General Education required 45 faculty members--25 professors, 9 GAs, and 11 adjuncts, thus using 84% of the faculty and approximately 45% of the available class loads.

The Study

Purpose

The intial purpose of the study was to see whether or not faculty agree on criteria for evaluating student writing, and to what extend would they give similar scoress even if their criteria do not agree. I make no value judgments concerning the quality of instructor or insruction. To a large extent, agreement on scores in general reflect standards even if explicit criteria do not. Particular disagreement is only troublesome when reliability must be maintained; this is no longer a given (cf. Moss).

Scope

Using 15 sample student papers written for the WCE, this study analyzes the scores obtained by current writing faculty using a four point holistic scale in addition to one 3x5 card's worth of comments. As the score did not reflect actual grades (which would have been helpful for a more complex study), the small survey size became problematic for statistical analysis; however, some tentative conclusions are possible.

Methodology

The study participants were obtained in order to reflect a representative sample of the current composition staff (this sample accounts for only 22.7% of the English faculty but reflects 46% of the number of current CRW teachers, even if participants were not currently teaching the course); each filled out a consent form and demographic information (Appendices A and B).

Discussion

While several variables might have affected scores (notably the prompt and writing mode), none were insurmountable for the study.

Criteria

Anecdotal comments. One of the assumptions of the WCE was that "all valid methods of evaluating writers' work are based on criteria of one kind or another, stated or assumed" (Report 15). Regarding current assumptions, Richard Straub and Ronald Lunsford indicate that "at this stage in the development of our discipline, we have no consensus as to what constitutes good writing" (12). Peter Elbow and Kathleen Yancey suggest that readers (exemplified by English as a discipline) are generally rewarded for divergent points of view rather that conformity (93-94). The raters in my study quickly demonstrated this lack of consensus which has a long research history (Diederich; Lunsford; Littlefield et al.; Spandel and Stiggins; Connors and Lunsford). While the scores were generally within a range, the anecdotal comments gave a brief glimpse of the criteria used and how the same essay showed positive and negative instances of the same criteria.

Since the study allowed raters only the front of a 3 x 5 card to record comments, the raters could only make global comments and could not even approach documenting every error. In fact, most raters only made approximately 4 comments per card (mean = 4.08, range of total comments = 20-101). Yet, anecdotal evidence from the holistic assessment reflected how the impressions of errors influence overall scores, in fact number and type of error was not always noted by the score (cf. Sloan; Straub and R. Lunsford). Broadly, though, content and form were relatively easy to distinguish without subjecting the comments to reliability analysis (cf. Appendix C, Table 8). While many of the comments were short words describing a feature of the writing, some involved advice given to the writer using "you." Some comments, such as redundant and "strays," could have been indicative of either form or content. Word count and

legibilty were classified as form. Table 2 gives a list of the kinds of comments comprising content and form.

Using the content and form distinction, I tabulated and categorized the anecdotal comments (see Appendix C, Table 8 for the breakdown by rater). Additionally, I classified comment types as either positive or negative (see Appendix C, Table 9 for the breakdown by rater). I gave raters the number of comments equal to the number of examples, but not including the general rule. I classified raters as having a propensity (60%) of either (form or content and negative or positive). As a group the majority of comments were classified as mostly on form and mostly negative (see Table 3; cf. Appendix I).

Statistical analysis. "On Average, People Will be Average."

Not surprisingly, scores in the current study reflected a strong tendency toward the middle range of scores. Table 4 demonstrates the central tendency. Raters not only demonstrated the tendency toward middle scores; they were not far off from the expected average of 2.5 (Table 5). As a group, GAs gave the lowest scores and adjuncts gave the highest scores, although the differences from the mean (.22 and .25) are not statistically significant. Professors were .03 lower than the mean, and they gave a wider range (SD = .85), although differences between groups were minimal.

The results, oddly enough, support both the research indicating the central tendency (Buley-Meissner 56) and the research indicating the possibility of papers receiving the range of score points (Deiderich). The majority of essays did (Table 6).

Statistical analysis (chi-square) on the subgroups revealed that of several factors the most significant was the relationship of scores to whether or not the rater made more content or form comments, followed by percentage of positive comments and teaching classification. Scoring

experience was less significant, and having a pedagogy class was not significant. Table 7 provides this analysis (cf. Appendix C for individual raters). Only the form/content reaches statistical significance at the generally accepted .05 level.

Given the diverse backgrounds and training of the current faculty, raters unsurprisingly demonstrated a variety of scores and anecdotal comments. The data sufficiently produced information on certain trends, writing assumptions, and criteria which will allow for preliminary evaluation of the use of diverse faculty in a service course as a part of the university writing program.

Evaluating First-Year Writing

Because of the many skills necessary for students to move successfully into academic discourse, CRW has added the reading component since its creation, and currently research and pretesting is ongoing to determine the feasibility of adding an oral communication component. Yet, whether the current system is broken or not remains an issue. Susan McLeod acknowledges the external "pressure to evaluate individual programs in order to demonstrate their effectiveness" (373). Currently one of the issues within the English Department at UNI is how to finance the cost of supporting general education classes such as CRW, and one proposal was to eliminate the class and return to a system which allowed students to demonstrate essay writing mastery with an exam similar to the former WCE. Before another exam takes the place of this class, several issues concerning the validity of the interpretations of the assessment should be addressed through research, specifically as this validity relates to backwash and planning concerns (cf. Moss 236). The old system was broken because of the backwash and concerns over validity of having one essay demonstrate the competence needed for the variety of skills in different academic or professional writing. McLeod further notes that research should

acknowledge the purpose and audience for whom the data is prepared before proposing alternatives (379-80). Pamela Moss suggests that program directors must also decide if the outcomes of the assessment are ethical; that is, do assessments produce harmful backwash by their very nature of being assessments (235-36)?

Thus, the nature of the institution and writing's place within it continues to evolve. Scholars as far back as Ross Jewell (cf. Haswell who acknowledges that writing requires constant practice to maintain learning 314-319) argue that writing classes do not improve student writing in the long run. The institution itself demonstrates how integrating teaching and evaluating writing remains difficult.[7] If teachers see writing as contextual, then students must write differently for each class. Maintaining consistency is difficult when current practice values student writing which reflects purpose, audience and voice, rather than general, impersonal, one-size-fits-all essays.

An underlying assumption of the WCE was that writing well on WCE translated into writing well for other tasks. This continues through on the assumptions behind first-year writing as service. Lange (qtd in Haswell 22-23) partially undermines this by finding that students wrote lower quality essays outside of their concurrent English class. How then are students able to apply learning and writing done in with one audience and purpose in mind, to other writing situations? How does this reflect the service mandate?

The Strengths and Weaknesses of a Class

When given the chance to create or implement a class such as CRW, how can teachers expect to continue the consistency expected of large-scale assessment, let alone demonstrate that the discipline or teacher understands what good writing is and how to perpetuate it? Lee Odell acknowledges some of the tensions raters or teachers might have, "We do seem to internalize a

lot of our assumptions and habits without conscious reflection." ("Introduction" 4). While most of the respondents (9, with 4 currently enrolled) mentioned having pedagogy classes, traditionally English teachers are far better than average in writing ability and may not have had training in composition. Whether teachers obtain a composing theory or not, Odell notes that

> we may have spent a lot of time embedded in contexts of practice that we may not want to perpetuate. . . . Moreover, to the extent that we received any writing instruction at all, there's a good chance it grew out of the practical stylist tradition that emphasized correctness, ignored the process of constructing meaning, and assumed that we should know what we want to say before we started to write. ("Introduction" 4)

Additionally, one of the issues central to the original assessment was that the scoring was done by professional (tenure-track) faculty. Since my study did not find significant differences between the subgroups, one might suggest that standards, or at least consistently applied standards, are not sacrificed in spite of the current practice which reflects a more diverse and possibly less experienced group of teachers. Lloyd Rieber notes that one of the assumptions behind the assessment of writing is that "most writing teachers would agree that the only way to evaluate students' writing ability is to evaluate sample of their writing" (15). But, Rieber acknowledges, "If you accept this notion, a major bottleneck in writing classes becomes the evaluation process" (15).[3] Susan Miller acknowledges that little has truly changed since Kitzhaber's critique in 1963 (11). Miller notes that Kitzhaber "characterized teachers of writing as graduate students and junior instructors whose status pleases administrations in need of cheap labor, pleases senior professors needing graduate students, and pleases graduate students who need work" (11). Miller further suggests that the assumptions of composition are seldom challenged by teachers who have no time (busy teaching and grading) for "self-reflection." Miller writes, "In a magnificent tautology, the practices that take our time are already validated, even and especially in their temporal demands, on the apparent 'need' for them" (15). Miller's

Appendix (pages 205-260) establishes Composition's lack of status in the English Department and reinforces UNI's ongoing struggles with handling the backlog of students.

Miller suggests the current system works for administration, faculty, and graduate students, but does the system work for students? My study invites a scientific metaphor--the (overused?) Heisenberg principle of uncertainty. The observer--administrators, teachers, researchers--can either see how a student is doing in a particular class (student grade) or see how students in classes are doing (class average), but cannot predict with any certainty particular grades or class averages: on average, people will be average. The total class average of all classes will demonstrate central tendency[4] but how particular students or instructors fit cannot be determined. So long as students are content with the average and range, they will not see the problem; when they compare workloads or criteria between classes, they might acknowledge the problem by avoiding certain teachers or classes. This global method of choosing classes by means of avoiding extremes works well for informed students who can take the time to pick and choose easier or harder instructors, or choose the learning environments they believe will best meet their needs. But, the class is designed for first-year students who come to the university usually already enrolled in a class, or who take the class during the second semester. Since the university cannot predict how that student will do (not that writing ability itself is predictable) nor fully demonstrate that all classes will be equal, the best the student can hope for is central tendency. My study did not attempt to evaluate learning outcomes, nor the financial reasons for one means of obtaining those outcomes, nor did it initially seek to propose any one alternative over the others, especially since many alternatives require similar assessments. My study suggests that 1) the current system produces scores (and by inference grades--just over a 3.0) which tend toward the average with few highs and fewer lows, 2) no standardization of criteria

exists among teachers of CRW, 3) we serve the university by promoting academic freedom for all teachers of first -year composition. As other research critique large-scale assessments as sacrificing autonomy to achieve standardization (Elbow), allowing diversity gives individual teachers the power to promote more contextual standards. Diversity within a community does foster maintaining at least the semblance of standards and who, if not teachers of writing, are best able to promote them.

I would like to thank my thesis committee (Scott Cawelti, John Swope, and Karen Tracey) for the impetus for completing this study. I also thank the teachers who participated in the study. Appreciation goes to my colleagues at the University of Louisville who asked the right questions concerning the relevance of the work to the topic at hand, although any incogruity is of my own. I thank the panel and the chair, a familiar face, for fellowship and advice.

Notes

[1] They probably did not envision Graduate Assistants or adjuncts still working on their MAs.

[2] Research indicates that institutions require different outcomes based on the nature of research and teaching. The role of the institution or even of writing itself is to promote strategies for teaching or research. Thus, do teachers and researchers see things differently? According to Marshall, "there remain the differences between them--differences which begin with contrasting institutional expectations and end with contrasting visions about what writing and the teaching of writing might be" (3) Marshall notes that research looks for the general while teachers focus on the particular (5).

[3] Rieber found that instead of having students, including graduate students and undergraduate English students, neither of whom had "a command of grammar, punctuation, and mechanics--who knew the rules, how to apply them, and how to explain them"serve as graders, paraprofessional editors (copy editors) were chosen to grade papers (16). These graders read and evaluated the papers twice (once for form and once for content) and were able to have a quick turn around time (Rieber 17). Reiber notes that this worked for six sections of fifty students (16), and included more writing than would have been practical with these numbers (17), and the graders were also able to consistently apply standard and effective tutor one hundred and fifty students a week (17).

[4] On an "A-F" scale, chances are more likely "A-C" will average toward a "B" since "D/F" students quickly withdraw. Additionally, grade inflation raises the expected, but averages can be predicted from previous years. For instance, in the Fall 1988, the GPA in CRW was 2.84 and 61% of students received a "B" or better; in the Spring of 1996, the GPAin CRW was 3.09 and 74% of the students received a grade of "B" or higher (Grade Distribution for 620:005).

TABLE 1

Faculty Responsibilities in General Education during Fall 1996

	Number of Sections (Number of Different Faculty Teaching the Course)				
	620:005	620:015/034	620:031	Hum/AmCiv	Total
Professors	15 (12)	5 (4)	14 (10)[a]	8 (7)	42 (25)[b]
Adjuncts	16 (9)				16 (9)
GAs	11 (11)				11 (11)
GAs (no teach)					2 (2)[c]
Total	42 (32)	5 (4)	14 (10)	8 (7)[c]	69 (45)[b]

[a] 3 additional sections are taught by non English faculty. [b] Faculty overlap on GE assignments
[c] Humanities accounted for 2 assignments in addition to 1 GA with unspecified duties. They were not figured in totals.

TABLE 2

Examples of Comment Types

Content	Form
Support seems to support a different point	transitions/continuity/flow (+/-)
selfish/shallow/transparent superficial	style/language (+/-)
logical/one-sided	not 500 words
sentimental/cheesy	fragment/run on, etc.
examples (+/-)	structure/organization (+/-)
generalizations	misspellings/homonym error
introduction (lead in)/conclusion (+/-)	colloquial/informal/scholarly use)
viewpoint	tone
idea organization	punctuation
difficult	5-paragraph mold (too obvious)
rambles	clear and concise
inadequate development	no proofreading
job/ not career	idiom/cliche
says nothing ("blah, blah, blah")	wordy
entertaining/personable	mixed constructions
excited about topic	meets assignment
states obvious	gendered language
point unclear, focus (+/-)	marginal mechanics
lot of B.S.	long paragraphs
blatant contradiction	purple prose
redundant ideas[a]	redundant (words)
logical progression (+/-)	points in intro and followed each
telling not showing	technical expression
Less certain	
strays/does not	didactic (use of "you")
"on task"	good context

[a] redundant was placed under form unless explicitly about ideas

TABLE 3
Response Type

Type	Responses	Number	Percentage
	Negative	581	63.3
	Positive	338	36.7
	Form	518	56.4
	Content	401	43.6

n = 919

TABLE 4
Central Tendency

Score Point	1	2	3	4
Totals	20	86	82	37
Percent	8.8	38.2	36.4	16.4

TABLE 5
Scores by Groups

	Scores		Score Point			
Group	AVG	SD ($\chi\sigma n$)	1	2	3	4
GAs	2.38	.84	10	34	23	8
Adjuncts	2.85	.82	5	17	37	16
Professors	2.57	.85	5	35	22	13
Totals	2.60	.74	20	86	82	37

TABLE 6
Range of Scores

Number of Essays given each score point: 8
(7, 24, 27, 28, 35, 39, 83, 113)
Also 4 essays received two 1s and two 4s.
(24, 28, 35, 113)
Only 5 essays were given half or more at one score point:

44	12 x2s
66	10 x2
73	9 x2
88	9 x3
105	8 x2 (also had 7x3)

TABLE 7
Two Column Analyses by Subgroups

Content or Form on Anecdotal Comments

Group	Score Point	1	2	3	4	Total
Content (n = 4)		8	26	21	5	60
Form (n = 8)		12	34	48	21	120
Neither (n = 3)		0	26	13	6	45

df = 6 chi-square 34.09 p<less than .0005

Positive/Negative on Anecdotal Comments (% is Positive)

Group	Score Points	1	2	3	4	Total
0-29.9 (n = 3)		6	22	14	4	45
30-39.9 (n = 7)		7	34	44	20	105
40-49.9 (n = 2)		7	12	8	3	30
50.0 + (n = 3)		0	18	16	11	45

df = 9 chi-square 22.49 .01>p>.005

Job Description[a]

Group	Score Points	1	2	3	4	Total
GAs (n = 5)		10	34	23	8	75
Adjuncts (n = 5)		5	17	37	16	75
Professors (n = 5)		5	35	22	13	75

df = 6 chi-square 17.45 .01>p>.005

Scoring Experience

Group	Score Points	1	2	3	4	Total
Yes (n = 5)		1	32	27	15	75
No (n = 10)		19	54	55	22	150

df = 3 chi-square 8.312 .05>p>.025

Pedagogy

Group	Score points	1	2	3	4	Total
Yes (n = 9)		14	48	50	23	135
No (n = 6)		6	38	32	14	90

df = 3 chi-square 1.55 p> .25

[a] The number of classes taught corresponded to job description.

APPENDIX A

RECRUITMENT LETTER

Memorandum
To: English Faculty
From: Scott Cawelti
Buzz Pounds
Date: 11 September 1996
Re: Participation in a Thesis Research Study

Dear Colleagues:

One of our MA students (Buzz Pounds) is conducting a research study for his Master's Thesis to ascertain how and by what criteria essays are scored by teachers assigned to teach 620:005. Since many faculty are periodically assigned to sections of this class, we would appreciate widespread participation.

The essays will be distributed in two batches around the first week of October, and we need to have them returned by the end of October. Participants will be asked to score 15 essays holistically and provide brief anecdotal information about what criteria was used to score each essay. We are asking participants to spend no more than 3 minutes reading each essay so that the total time commitment should not exceed an hour to an hour and fifteen minutes.

Please indicate below whether you would be willing to participate.

Name_____
_____ Yes, I will participate.

Thank you in advance for your cooperation

APPENDIX B

DEMOGRAPHIC INFORMATION

Questionnaire

Please take a minute to fill in the appropriate demographic information.

Position:

_____ Professor
_____ Associate Professor
_____ Assistant Professor
_____ Adjunct
_____ Teaching Assistant

Educational Level (Highest):

_____ PhD
_____ MA (English)
_____ MA (Other)
_____ More than 9 Graduate hours credit
_____ BA

Number of classes taught in each of these Beginning or Intermediate Prose Writing course, either here or an equivalent class elsewhere:

_____ 620:005 (College Reading and Writing)
_____ 620:015 (Expository Writing)
_____ 620:034 (Critical Writing about Literature)
_____ 620:103 (Personal Essay)
_____ 620:104 (Argument and Persuasion)

Have you ever taken a writing methodology or pedagogy class?

_____ Yes, if so, how recently? _____
_____ No

Have you ever participated in a large-scale evaluation of writing, for example scoring the former UNI writing competency exam?

_____ Yes
_____ No

Thank you

APPENDIX C

RATERS BY CLASSIFICATION

TABLE 8
Raters By Comment Category

Rater	Total Comments	Content	Form	Percent	Score Average
Content Raters					
5	48	38	10	80.8	2.0
15	101	74	27	73.2	2.0
11	89	59	30	66.2	2.4
8	57	35	22	61.45	3.0
Form Raters					
1	75	13	62	88.0	3.0
13	54	7	47	87.0	3.1
9	52	12	40	76.9	2.6
3	52	16	36	69.2	1.9
6	63	20	43	68.2	3.0
7	52	17	35	67.3	2.5
10	53	19	34	64.1	3.1
4	78	30	48	61.5	2.6
Neither					
14	82	33	49	59.7 (F)	2.9
2	43	18	25	58.1 (F)	2.5
12	40	20	20	50.0	2.3

Rater 14: note--does not include a comment on handwriting nor a comment on the assignment itself as possible reasons for ambiguity

Rater 11: note--the Form comments included "Organized, but mechanical" listed on 10 responses and counted as both a positive and a negative comment.

Rater 13: note--the Form comments included "Meets Assignment Guidelines" listed on 9 responses and counted as a positive comment.

TABLE 9
Raters by Comment Type

	Content		Form		Percent +
Rater	Positive	Negative	Positive	Negative	
13	2	5	33	14	64.8
11	35	24	16	14	57.5
8	21	14	9	13	52.6
4	18	12	20	28	48.7
3	6	10	15	21	40.3
1	2	11	27	35	38.6
9	3	9	17	23	38.4
2	6	12	10	19	37.2
6	14	6	9	34	36.5
5	10	28	7	3	35.4
14	12	23	17	32	32.9
10	9	10	8	26	32.0
15	18	56	3	24	20.7
7	6	11	4	31	19.2
12	0	10	0	10	00.0

Scorer 14: note--does not include a comment on handwriting nor a comment on the assignment itself as possible reasons for ambiguity

Scorer 11: note--the Form comments included "Organized, but mechanical" listed on 10 responses and counted as both a positive and a negative comment.

Scorer 13: note--the Form comments included "Meets Assignment Guidelines" listed on 9 responses and counted as a positive comment.

BIBLIOGRAPHY

Baker, Eva L. "Critical Validity Issues in the Methodology of Higher Education Assessment." Assessing the Outcomes of Higher Education: Proceedings of the 1986 (47th) ETS Invitational Conference. Princeton: ETS, 1987. 39-46.

Buley-Meissner, M. L. "Reading Without Seeing: The Process of Holistic Scoring." Writing on the Edge 4 (1992): 51-65.

Chandler, John W. "The Why, What, and Who of Assessment: The College Perspective." Assessing the Outcomes of Higher Education: Proceedings of the 1986 (47th) ETS Invitational Conference. Princeton: ETS, 1987. 11-18.

Connors, Robert J., and Andrea A. Lunsford. "Frequency of Formal Errors in Current College Writing, or Ma and Pa Kettle do Research." College Composition and Communication 39 (1988): 395-409.

Diederich, Paul B. Measuring Growth in English. Urbana, IL: NCTE, 1974.

Eblen, Mac. Personal Interview. 18 February 1997.

Elbow, Peter, and Kathleen Blake Yancey. "On the Nature of Holistic Scoring: An Inquiry Composed on EMail." Assessing Writing 1.1 (1994): 91-107.

Grade Distribution for 620:005. Cedar Falls: UNI Registrar's Office, nd [1996].

Haswell, Richard. Gaining Ground in College Writing: Tales of Development and Interpretation. SMU Studies in Composition and Rhetoric. Dallas: SMU P, 1991.

Hughes, Arthur. Testing for Language Teachers. Cambridge (UK): Cambridge U P, 1989.

Jewell, Ross M. The Effectiveness of College-Level Instruction in Freshman Composition. Washington: U.S. Department of Health, Education, and Welfare, 1969.

Lederman, Marie Jean. "Why Test?" Writing Assessment: Issues and Strategies. Ed. Karen L. Greenberg, Harvey S. Wiener, and Richard A. Donovan. Longman Ser. in College Composition and Communication. New York: Longman, 1986. 35-43.

Littlefield, John, et al. Analyzing Written Comments by Performance Raters. Paper presented at the Annual Meeting of the American Educational Research Association, San Franscico, CA, April 20, 1992. Washington: ERIC, 1992.

Lunsford, Andrea A. "The Past--and Future--of Writing Assessment." Writing Assessment: Issues and Strategies. Ed. Karen L. Greenberg, Harvey S. Wiener, and Richard A. Donovan. Longman Ser. in College Composition and Communication. New York: Longman, 1986. 1-12.

McLeod, Susan H. "Evaluating Writing Programs: Paradigms, Problems, Possibilities." Journal of Advanced Composition 12 (1992): 373-82.

Miller, Susan. Textual Carnivals: The Politics of Composition. Carbondale, IL: SIU P, 1991.

Moss, Pamela A. "Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment." Journal of Educational Research 62 (Fall 1992): 229-58.

Odell, Lee. "Introduction." Theory and Practice in the Teaching of Writing: Rethinking the Discipline. Ed. Lee Odell. Carbondale, IL: SIU, 1993. 1-8.

Reynolds, W. Ann. "Higher Learning in America: Aims and Realities." Assessing the Outcomes of Higher Education: Proceedings of the 1986 (47th) ETS Invitational Conference. Princeton: ETS, 1987. 1-10.

Rieber, Lloyd. "Paraprofessional Assessment of Students' Writing." College Teaching 41.1 (Winter 1993): 15-18.

Sloan, Gary. "Frequency of Errors in Essays by College Freshmen and by Professional Writers." College Composition and Communication 41 (1990): 299-308.

Spandel, Vicki, and Richard J. Stiggins. Creating Writers: Linking Assessment and Writing Instruction. New York: Longman, 1990.

Straub, Richard, and Ronald F. Lunsford. Twelve Readers Reading: Responding to College Student Writing. Cresskill, NY: Hampton P, 1995.

University of Northern Iowa Faculty Senate. Minutes. 24 January, 1977.

---. Minutes. 11 May 1981.

---. Minutes. 16 November, 1987.

University of Northern Iowa Department of English Language and Literature. The University of Northern Iowa Writing Examination Program [Report]. January, 1980.

---. Fall 1996 Schedule of Classes. Corrected. 21 August 1996.